Title

Targeted Switchgrass BAC Library Screening and Sequence Analysis Identifies Predicted Biomass and Stress Response-Related Genes

Journal

Bioenergy Research, 9(1)

ISSN

1939-1234

Authors

Sharma, MK
Sharma, R
Cao, P
et al.

Publication Date

2016-03-01

DOI

10.1007/s12155-015-9667-1

Peer reviewed

Targeted switchgrass BAC library screening and sequence analysis identifies predicted biomass-related and stress response genes

Manoj K Sharma[1,2,3,#], Rita Sharma[1,2,4,#], Peijian Cao[5], Mitch Harkenrider[1], Jerry Jenkins[6,7], Jane Grimwood[6,7] Jeremy Schmutz[6,7], Michael K. Udvardi[8], Daniel S. Rokhsar[7,9], Pamela C. Ronald[1,2,*]

[1]Department of Plant Pathology and the Genome Center, University of California, Davis, California, United States of America; [2]Joint BioEnergy Institute, Emeryville, California, United States of America; [3]Present Address: Department of Biotechnology, Jawaharlal Nehru University, New Delhi, India; [4]Present Address: Department of Life Sciences, Jawaharlal Nehru University, New Delhi, India; [5]China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute, Zhengzhou, China; [6]HudsonAlpha Institute of Biotechnology, Huntsville, Alabama, United States of America; [7]United States Department of Energy Joint Genome Institute, Walnut Creek, California, United States of America; [8]Plant Biology Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, Oklahoma, United States of America; [9]University of California, Berkeley, California, United States of America.

[*]Corresponding Author
Email: pcronald@ucdavis.edu

[#]Equal Contribution

Abstract

To identify switchgrass homologs of rice genes predicted to control biomass-related traits, we screened 96,000 clones from two switchgrass BAC libraries. Full-length sequencing of 311 selected BAC clones revealed sequence for ~3.2% (51.7 Mb) of the switchgrass genome, coding for 3948 genes (4217 gene models). The targeted cell wall and stress response-related genes, including 350 kinase, 203 glycosyltransferase, 109 glycoside hydrolase and 33 ethylene responsive transcription factor genes, are particularly enriched among the annotated genes. Using a phylogenomic and structural approach, we demonstrate that most of the switchgrass glycosyltransferase 2 (GT2) gene subfamily members cluster into previously established subfamilies. However, five of these genes, together with predicted orthologs from maize, *Sorghum* and poplar form a separate clade, that are absent in rice and *Arabidopsis*. Comparative expression analysis using microarrays and qPCR-based assays revealed high correlation in expression profiles of GT2 homologs in rice and switchgrass suggesting that rice genes are predictive of the switchgrass gene functions.

Keywords: BAC library, biofuel, cellulose synthase, glycoside hydrolase, kinase, screening, switchgrass

Introduction

Switchgrass is a candidate feedstock crop for biofuel production [1]. Due to its low input requirements and adaptability to a wide range of growing conditions, switchgrass stands can be grown on marginal lands not suitable for production of most food and feed crops [2-4]. Yet unlike sugars, starches and oils from first generation biofuel crops such as corn and sorghum, the source of energy in switchgrass is the less tractable lignocellulosic biomass [5-8]. Another point is that like any crop grown on large scales for extended periods of time, switchgrass stands will be exposed to diverse biotic and abiotic stresses. For these reasons, research directed at optimizing switchgrass for biofuel production has focused on two fronts. First, the saccharification efficiency must be improved to make energy extraction cost effective [9,10]. Secondly, varieties that can withstand diverse stresses must be developed to help stands thrive on marginal soils and in face of a variety of pests and

diseases [11].

Genetic studies on switchgrass have been limited due to its complex genomic structure and perennial, outcrossing habit [12,13]. For this reason, researchers have utilized genetic resources from rice, a model grass species, to advance understanding of switchgrass biology. For example, rice experimental studies using phylogenomic, transcriptomic and other genetic analyses have identified glycosyltransferase (GT), glycoside hydrolase (GH), kinase and ethylene responsive transcription factor (ERF) gene families and demonstrated an important role for these genes in cell wall biosynthesis, and modification and tolerance to stress [14-19] [20-23]. In this study, we sought to identify switchgrass homologs of rice genes belonging to these four (GT, GH, kinase and ERF) gene families.

Leveraging the switchgrass BAC library resources available, we organized 96,000 clones from two libraries into pools and superpools and established a qPCR-based screening workflow to identify BAC clones carrying the targeted genes. Using this screen, we selected and sequenced 311 BAC clones corresponding to 51.7 Mb of the switchgrass genome. We identified 3948 unique genes from this sequence, 695 belonging to the targeted families (GT, GH, kinase and ERF). Further, we compared gene organization between the sequenced switchgrass BAC clones and the corresponding regions in the rice genome, finding significant colinearity. Lastly, we compiled a list of 65 switchgrass orthologs of rice genes having demonstrated roles in bioenergy-relevant traits using genetic and biochemical approaches. This list includes 14 cellulose synthase and cellulose synthase-like genes, a glycoside hydrolase gene, two lignin biosynthetic genes, four genes regulating flowering time, and 32 genes central to the plant biotic and abiotic stress responses.

Results and Discussion

In this study, we sought to identify full-length sequences of switchgrass genes predicted to control cell wall biosynthesis and modification and stress-response related traits.

BAC library pooling and superpooling enables quick and accurate screening of large libraries

BAC libraries are a valuable resource for target gene identification, however, conventional screening of large libraries is an expensive and time-consuming process.

Pooling BAC clones is an efficient strategy to minimize false positive hits and reduce the effort required for screening a large number of clones [24,25]. We used seven-plate matrix pool & superpool technology developed by Amplicon Express (http://ampliconexpress.com/) to organize 96,000 clones from two switchgrass BAC libraries [26]. Taking into account the size of the switchgrass genome (1600 Mb) and average BAC length (144 and 110 Kb) in each library, we estimate ~4X and 3X coverage of the genome in respective pools and superpools (P&SPs). Based on the Clarke-Carbon's formula [27], there is a 99.9% probability of finding a specific sequence in our pooled collection. Our collection comprises 18 superpools (SPs) and 23 matrix pools (MPs). Each SP contains pooled DNA from 2688 BAC clones. The MPs include five plate pools (PP), eight row pools (RP) and ten column pools (CP) [24,28]. The screening scheme used in this study is comprised of two rounds of PCR. The first round involves 20 reactions to screen the 18 SPs along with positive and negative controls. The second round screens 23 MPs and controls. Therefore, a set of 45 PCR reactions (20 with SPs and 25 with MPs) is sufficient to find a BAC clone carrying the gene-of-interest (Supplementary figure 1; see methods for details). Instead of conventional PCR followed by gel electrophoresis, we used real-time qPCR and melt curve analysis to screen the matrices. [29].

Cell wall and stress response-related genes were targeted for BAC library screening

Over the past two decades, researchers have identified genes regulating biomass quality, sugar yields, abiotic stress tolerance and defense response in grasses using rice as an experimental model. Here, we targeted switchgrass homologs of rice genes belonging to GT, GH, kinase and ERF families that have roles in cell wall composition, saccharification efficiency and stress tolerance [14,15,17,20,21]. In rice, a total of 2654 annotated genes belong to these families (609 GT, 437 GH, 1467 Kinase and 145 ERFs). To narrow this list of gene targets, we focused on those with known functions in rice or genes exhibiting high expression in above ground organs. We also hunted for several defense-response related genes comprising the rice stress response interactome elucidated in our laboratory [30].

Several studies have reported clustering of gene family members in plant genomes as well as marked colinearity between grass genomes [26,31]. Therefore, to avoid selecting multiple BACs from the same genomic region, we localized our target

genes on the rice chromosomes. Wherever, our target genes were clustered on rice chromosomes, only one set of primers was designed from a 200 Kb block, thereby avoiding the selection of multiple BAC clones representing identical genomic regions (Figure 1). Because qPCR primers targeted to rice gene sequences only had a 10% success rate in amplifying the switchgrass ortholog, we generated switchgrass-specific primers for 427 genes based on sequences in public repositories. For this purpose we used unique transcript sequences (http://switchgrassgenomics.noble.org/) and EST sequences of switchgrass (http://wheat.pw.usda.gov/panicum/blast/, accessed March 2010). The efficiency and specificity of the primers were tested using switchgrass genomic DNA as a template followed by melt curve analysis. Of the 427 primer pairs, 106 either did not amplify or gave multiple amplification peaks. The remaining 321 primer pairs were used for screening the BAC collection. During the screen 55 of the 321 primer pairs amplified sequences from more than two matrix pools therefore, a specific BAC address could not be determined. Amplification signals from multiple pools may be due to the presence of multiple copies of the genes-of-interest in the switchgrass genome, high sequence similarity among members of a gene family or high enrichment of the targeted genomic fragment in the BAC libraries. Using this screen, a total of 266 BAC clones carrying targeted genes were identified.

To identify more BACs carrying the targeted genes-of-interest, we computationally screened 330,297 BAC-end sequences (BESs) derived from the switchgrass BAC libraries for our targeted genes [26]. Forty-five of these BESs contained partial sequences for our genes-of-interest and were chosen for full-length sequencing. Ultimately, a total of 311 BACs were sequenced to full-length using Sanger sequencing.

Analysis of 311 BAC clone sequences provided information for 51.7 Mb switchgrass genome encoding 4217 predicted transcripts

The sequence from 311 BAC clones comprises 51.7 Mb that represents ~3.2% of the switchgrass genome. Insert size of the selected BAC clones ranged from 46 to 357 Kb with an average insert size of 166.2 Kb (Supplementary table 1). The GC content of the sequenced BACs ranged from 43 to 53.5%, with an average of 47.3% across the 51.7 Mb of sequence analyzed (Supplementary table 1).

A total of 4217 transcripts (gene models) corresponding to 3948 genes (Supplementary table 2) were annotated from the sequence obtained in this study.

Full-length genomic, cDNA and protein sequences as well as the annotation information file (.gff file) of the predicted genes are provided in supplementary files 1 to 4. Although the gene density varies from BAC to BAC, the average gene density is estimated to be one gene per 13.1 Kb (Supplementary table 1). For 68% of the predicted genes, we could identify at least one of the UTRs, whereas, 109 genes located at the BAC ends were incomplete. The average number of exons and introns per mRNA is 4.6 and 3.6 respectively (Supplementary table 1).

When analyzed for SSRs and known plant repeat elements, we identified 10,334 SSRs (Supplementary table 3). The average density of SSRs was one SSR per 5 Kb of the sequence analyzed. About 52% of SSRs are trimers and ~76% of these are GC rich. 30,810 repeat elements were identified corresponding to ~28% of the sequence analyzed (Supplementary table 4). 509 small RNAs, 131 satellites and 2094 low complexity regions were identified (Supplementary table 4).

Targeted BAC screening and sequence analysis revealed full-length sequences of 695 genes belonging to GT, GH, Kinase and ERF families

Alamo switchgrass and rice differ in their genome sizes (1600 vs 373 Mb), ploidy levels (tetraploid vs diploid), breeding systems (outcrossing vs inbreeding) as well as life cycles (perennial vs annual). Despite dramatic differences in habit and genetic organization, rice and switchgrass have significant synteny and colinearity [26,31,32].

In this study, we leveraged the conservation between the rice and switchgrass genomes to identify switchgrass homologs of rice genes belonging to the GT, GH, kinase and ERF families. Out of the 311 BAC clones analyzed here, 274 contained genes-of-interest (88%). As many of the BACs were selected to target clusters of genes-of-interest, 185 BACs contained more than one gene-of-interest. About 24 BACs did not contain the specific targeted gene but did contain another member of the same family. The high degree of sequence similarity among members of a gene family might explain this off-target identification.

Gene Ontology analysis highlighted the high representation of the cell wall and stress-related genes among the annotated genes (Figure 2). Out of 3948 unique genes annotated in this study, 695 belong to GT, GH, kinase and ERF gene families. Among these, 350 encode kinases, 203 GT, 109 GH and 33 ERF family genes. Many of these genes are predicted to exhibit alternative splicing and therefore, correspond to 374

Kinase, 214 GT, 123 GH and 38 ERF gene models. The conserved domains in the translated genes were confirmed by Pfam domain analysis. Further information about the respective subfamilies has been provided in supplementary table 5.

Comparative analysis of switchgrass genes with genes in other plant genomes

We used discontinuous Mega BLAST (designed to find long alignments between similar but not identical sequences for cross-species comparisons) to determine the homology of switchgrass genes annotated in this study with other species including rice, *Sorghum*, foxtail millet, maize, *Brachypodium* and *Arabidopsis*. Out of 3948 loci annotated from switchgrass, we found corresponding homologs of 3770 (95.5%) in rice, 3619 (91.7%) in foxtail millet, 3611 (91.5%) in *Sorghum*, 3575 (90.6%) in maize, 3546 (89.8) in *Brachypodium* and 3128 (79.2%) genes in *Arabidopsis* (Supplementary Table 5). In several instances, two or more switchgrass genes had homology to the same gene in other grass species. These cases are likely due to multiple alleles in the polyploid switchgrass genome associated with a single locus in diploid relatives.

A total of 118 genes from the switchgrass sequences we analyzed did not show significant homology with genes from any of the other genomes used in this study. Five of these genes have Pfam domains corresponding to CCHC zinc finger, ribosomal L39, BED zinc finger, Tify and UBN2 domains. Thirty-three of these genes are located at the BAC ends. Out of these, fifteen are smaller than 300 nt in length and therefore, might be partial genes.

We also compared genomic organizations between the switchgrass BAC clones characterized and the corresponding regions in the rice genome. Out of 274 BAC clones containing genes-of-interest, 233 share significant colinearity and synteny with rice (Figure 1). The remaining 41 BACs either had too few genes to be included in the analysis or did not share significant synteny with rice.

Comparative phylogenetic analysis of GT2 family genes in switchgrass

The cell wall composition in grasses is quite different compared with the dicots [33,34]. Some of these differences are apparent from the evolutionary expansion and diversification of GT2 family cellulose synthase and cellulose synthase-like proteins in plants [34-37]. To investigate these relationships further and track the molecular evolution of GT2 family in switchgrass, we performed a comparative phylogenomic

analysis of GT2 family proteins from six grass and two eudicot genomes (Figure 3; Supplementary figure 2). Out of 87 GT2 family genes identified from switchgrass, 11 were <300 nt in length and therefore, not used for phylogenetic construction. A total of 506 GT2 proteins including 118 from maize, 76 from switchgrass, 64 from poplar, 55 from foxtail millet, 54 from *Sorghum*, 50 from rice, 47 from *Brachypodium* and 42 from *Arabidopsis* were used for phylogenetic analysis (Supplementary File 5).

Previous phylogenetic classification of GT2 proteins in plants has categorized this family into eleven major functional and evolutionary lineages including cellulose synthases (CesA), cellulose synthase-like groups (CslA, B, C, D, E, F, G, H and J) and dolichyl-phosphate beta-glucosyltransferases [37,38]. CesAs play central roles in cellulose biosynthesis. CslA genes encode mannan synthases, CslC are xyloglucan synthases, and CslF and H comprise mixed-linkage glucan synthases. The precise functions of other Csl families are not yet known, but they also have predicted roles in the synthesis of cell wall polysaccharides. Eleven of the switchgrass proteins, used for phylogenetic analysis, grouped with cellulose synthases, five with CslA, one with CslC, four with CslD, five with CslE and twenty-five with CslF proteins. As reported before, CslB and G groups were not represented in any of the grass genomes including switchgrass. We did not identify any switchgrass proteins representing CslH and CslJ clades in our study. However, nine gene models clustered with dolichyl-phosphate beta-glucosyltransferase proteins.

In several clades, proteins split into eudicot and grass-specific subgroups due to lineage-specific divergence in their gene structures (Supplementary Figure 2). In addition, we identified a separate clade with representative members from maize, *Sorghum*, poplar, and five switchgrass proteins that do not include proteins from the model systems, rice and *Arabidopsis* (Figure 3). Though these proteins showed ambiguous clustering between different phylogenetic trees, this clade likely represents highly-diverged, lineage-specific sequences.

To further strengthen our analyses, we combined phylogenetic grouping with conserved motif identification using Multiple Em for Motif Elicitation (MEME) tool ver. 4.9.0 (Supplementary Figure 3). Consensus sequences of the 20 conserved motifs, identified from GT2 genes are given in supplementary table 7. While motifs 2, 10, 11, 12, 17, 18 and 19 represent the conserved domain of cellulose synthases, all other motifs represent the conserved domain (PLN02893) of cellulose synthase-like proteins. Although the GT2 domain is present in all the groups based on MEME motif

analysis, we noticed sequence divergence in the GT2 domain between different GT2 subfamilies. As expected, most of the switchgrass genes had a similar domain distribution as the other grass genomes. Twenty-nine GT2 proteins of switchgrass, annotated as the first or last protein on a BAC clone, lacked some of the motifs, likely due to incomplete protein sequence. However, as illustrated in supplementary figure 3, several switchgrass GT2 proteins mostly belonging to the CesA clade and the new group identified in this study lacked the motifs identified from other genomes indicating the loss of a structural domain. The functional significance of these differences is unknown.

Microarray-based comparative expression analysis of cellulose synthase and cellulose synthase-like genes in rice and switchgrass

To determine if the expression patterns of CesA and Csl genes are conserved between rice and switchgrass, we analyzed publically available microarray-based data for both species [39,40]. The expression data for rice CesA and Csl genes from six organs (root, leaf blade, leaf sheath, stem and two stages of inflorescence development) were downloaded from the meta-data portal of the Rice Oligonucleotide Array Database (http://www.ricearray.org/expression/meta_analysis.shtml) [39]. Similarly, the expression data for switchgrass homologs in comparable stages of development was downloaded from the Switchgrass Functional Genomics Server at The Noble Foundation (http://switchgrassgenomics.noble.org/). Switchgrass homologs of fourteen rice genes are represented on the arrays, and four of these fourteen genes have two or more homologous sequences from switchgrass. Eight of the homologous gene pairs including *CesA1*, *CesA2*, *CesA9*, *CslA2*, *CslA11*, *CslC1*, *CslD4*, and *CslE6* of rice showed a Pearson correlation coefficient ≥0.7 with their switchgrass orthologs indicating conservation in expression and possible conserved function (Figure 4). On the other hand, six gene pairs including rice *CesA5*, *CesA6*, *CslC7*, *CslE2*, *CslF2* and *CslF8* did not show significant correlation with their switchgrass orthologs implying functional divergence.

Organ-specific expression profiling of GT2 genes in switchgrass using qPCR

To further deduce gene functions, we selected 24 genes representing different subfamilies of the GT2 family and checked their transcript accumulation patterns in six organs including root, panicle, mature leaf, young leaf, internode and node using

quantitative real-time PCR (Figure 5). Five genes encoding cellulose synthases (CesA) are preferentially expressed in panicle and stem tissues conforming to their expected role in secondary cell wall biosynthesis. Members of CslA group exhibit varied expression patterns with tissue-specific enrichment for different genes. For instance, 12694.m00008 transcripts are enriched in panicles and leaves, 13177.m00005 transcripts accumulate in roots, whereas, transcripts of 13177.m00004 were mainly detected in stems indicating functional diversity. CslD genes express in roots, panicles and stems. The CslE family genes are particularly enriched in leaf tissues with 15659.m00001 and 12733.m00006 expressing specifically in leaves while 13230.m00005 transcripts were also detected in roots and panicles. Among CslF genes encoding grass-specific mixed-linkage glucan synthases, two expressed in roots and panicles, whereas, one gene 15655.m00003 was specifically detected in stem internodes. Of the remaining three genes 15647.m00005 and 13175.m00013 encoding dolichyl-phosphate beta-glucosyltransferase express highly in leaf and stem tissues, respectively. Another gene, 13154.m00010, not belonging to any of the aforementioned groups, is also strongly expressed in stem tissues. Each selected gene model here is transcribed in at least one of the organs analyzed suggesting functional activity specific to that organ and developmental stage.

Thirteen of the 24 genes analyzed by qPCR were also represented on the switchgrass microarray. Most of these genes had comparable expression profiles between qPCR and microarray experiments in the stages analyzed except in roots. Three CesA genes (12060.m00014, 12061.m00004 and 12696.m00012) that were detected at very low levels in the mature roots in the qPCR analysis, express at high levels in root tissue used in the microarray experiments. This discrepancy between microarray and qPCR-based expression in roots may be due to the different developmental stages at which the samples were collected: mature roots for qPCR and elongation stage roots for microarray. These results corroborate the developmental-stage specific expression known to be a regulatory feature of cellulose synthases.

Bioenergy-relevant switchgrass genes with characterized rice orthologs

Because orthologous genes trace back to a common ancestor and are predicted to carry out biologically equivalent functions, orthology is a very powerful tool for transferring the functional information from experimentally-characterized genes in

model species to non-model organisms [41]. In this study, we have made an attempt to identify the switchgrass orthologs of rice genes with demonstrated functions in bioenergy-relevant traits using genetic and biochemical analysis. Out of the 3948 genes annotated in this study, rice orthologs of 65 genes have been previously characterized for their prominent roles in cell wall biosynthesis and stress response (Table 1). These include *CesA1*, *CesA2*, *CesA3*, *CesA5*, *CesA6*, *CesA8* and *CesA9* encoding cellulose synthases [35,42-44], *CslD4* involved in synthesis of hemicellulose [45,46], *CslF2*, *CslF3*, *CslF4*, *CslF6*, *CslF8* and *CslF9* playing central role in biosynthesis of mixed-linkage glucans [47][48]. We have also identified the switchgrass homolog of *OsGH9B5* that regulates cellulose biosynthesis and crystallinity [49] as well as *OsOMT1* (Caffeic acid O-methyltransferase) and *CAD2* (Cinnamyl alcohol dehydrogenase 2), genes involved in lignin biosynthesis [50-53]. We also identified switchgrass homologs of several other genes associated with saccharification yield including *OsSUT2*, a sucrose transporter [54], *IRX10-L (Irregular Xylem 10-L)* and *IRX14 (Irregular Xylem 14*) regulating xylan biosynthesis [55,56] and two genes, *OsBC1L4 (Oryza sativa brittle culm 1 like 4*) and *bc10* (*brittle culm 10*), regulating mechanical strength [57,58] (Table 1). Rice homologs of four of the targeted genes regulate flowering time (Table 1). Because, earlier studies have demonstrated increase in biomass by delayed flowering, these genes have potential to increase biomass yields by delaying reproductive growth in bioenergy grasses [59,60].

Earlier, we had reported construction of a stress response interactome around three rice proteins that control biotic and abiotic stress responses [30]. Subsequently, most of these interactions were found to be conserved in wheat indicating that this network has promise to advance the knowledge of stress responses in other grasses as well [61]. Also, many of the proteins comprising this interactome govern crosstalk between biotic and abiotic stresses and are promising targets for engineering broad-spectrum stress tolerance [16]. We have identified the switchgrass homologs of 13 genes from this interactome. These include *OsMPK6*, *OsMPK8*, *XB15*, *OsWAK25*, *SnRK1A*, *WRKY13*, *NH2*, *XB2IP-1*, *SAB4*, *SAB8*, *SAB9*, *SAB10* and *SCB3* (Table 1). Further analysis would reveal if they have conserved functions in switchgrass as well.

Finally, the switchgrass genome analysis reported here provides a foundation for detailed characterization, breeding and engineering of genes that regulate bioenergy-related traits in switchgrass.

Methods

Gene selection and primer designing

The meta-profiles of GT, GH, kinase and ERF family genes were downloaded from Rice Oligonucleotide Array Database (http://www.ricearray.org/expression/meta_analysis.shtml). Those exhibiting higher accumulation in above ground organs or implicated in cell wall biosynthesis/stress response, based on previous literature, were prioritized for screening.

We used nucleotide sequences of prioritized rice genes to extract corresponding assembled switchgrass unique transcript sequences (http://switchgrassgenomics.noble.org/; [40] and EST sequences of switchgrass from public repositories (http://wheat.pw.usda.gov/panicum/blast/, accessed March 2010) using blast search tool. The corresponding switchgrass sequences were BLAST searched in the Rice Genome Annotation Project Database (http://rice.plantbiology.msu.edu/analyses_search_blast.shtml) to identify unique regions for primer design. We used Beacon Designer 7.51 to design qPCR primers with default parameters specific for the SYBR® Green assay. If default parameters did not suggest potentially good primers, the annealing temperature was adjusted ($\pm 3^0$C) to find an optimal primer pair. We tested qPCR primers with genomic DNA (gDNA) of switchgrass and performed melt curve analysis to confirm the specificity of the primer pair. Primer pairs that produced specific amplicons with gDNA were used for screening.

Construction and screening of Pools and Superpools

Seven-plate Pool and Superpool (P&SP) system of Amplicon express (http://ampliconexpress.com) was used to identify BAC clones containing genes of interest. For each library, about 48,000 independently grown and then separately pooled clones were organized into seven-plate matrix pools and superpools. Each set comprised of 18 SPs with individual SP representing BAC DNA from 2688 clones. Each SP was divided into 23 matrix pools that comprised of five plate pools, eight row pools and ten column pools.

We used gene-specific qPCR primers to screen SPs and MPs as described (http://ampliconexpress.com/products-services/screening-services/pools-and-superpools) followed by melt curve analysis. Briefly, two rounds of qPCR were performed to identify BAC clones containing genes-of-interest. In round I, twenty qPCR reactions were performed on 18 SPs plus controls. Since the positive control nucleotide sequence is represented by single copy, SP giving amplification signal (Ct value) comparable to the positive control was selected for round II screening. In round II qPCR, 25 reactions were performed on 23 MPs plus controls. We also analyzed the melting curve profiles of the amplicons during both rounds of screening to confirm amplicon specificity. An online tool from Amplicon Express (http://puffer.ampliconexpress.com/) was used to analyze qPCR results and identify the selected BAC address in the library. When an accurate BAC address could not be determined from the first SP chosen, we selected another SP with a higher amplification signal in round I and repeated the second round of qPCR.

All the qPCR reactions were performed with SsoFast™ EvaGreen® mix using CFX96™ Real-Time PCR Detection System (Bio-Rad Laboratories, Inc. USA). The reaction conditions were as follows: 95^0C for 30 s followed by 40 cycles at 95^0C for 3 s and 54-60^0C for 3 s. Temperature-dependent dissociation characteristics of DNA amplicons were recorded at ramp from 65^0C to 95^0C, raising temperature by 0.5^0C at each step, pausing for 5 s followed by plate reading.

Full-length BAC sequencing and gene annotation

BAC clones were sequenced to full-length at the HudsonAlpha Institute of Biotechnology (www.hudsonalpha.org) by Sanger's method as described before [26]. We used Mreps (http://bioinfo.lifl.fr/mreps/) to screen for simple sequence repeats with the following parameters: 1-3 nt repeats at least 12 nt in length and 4-6 nt repeats with at least 4 unit repetition [62]. The known repeat elements were identified using RepeatMasker 4.0.2 (http://repeatmasker.org), [63,64] and AB-BLAST v3 based on the Viridiplantae section of the RepBase repeat database (release 20110419)(http://blast.advbiocomp.com/). All identified repeat elements were masked to produce high-quality genomic sequences. Gene models were identified using GenomeScan (http://genes.mit.edu/genomescan.html) from masked full-length BAC sequences. Further, PASA (Program to Assemble Spliced Alignments) (http://pasa.sourceforge.net/) and NCBI EST sequences were used to update

GenomeScan predictions. BLAST analysis in rice and *Arabidopsis* databases (http://rice.plantbiology.msu.edu/, http://www.arabidopsis.org/) and Pfam domain analysis (http://pfam.janelia.org/search) were performed to identify conserved domains.

Identification of homologs

Genomic sequences of rice, *Sorghum*, maize, foxtail millet and *Brachypodium* were downloaded from Phytozome v6.0 (http://www.phytozome.net/). The genomic sequence of *Arabidopsis* was downloaded from TAIR database V9 (http://www.arabidopsis.org/). Discontiguous Mega BLAST with a cutoff of $1e^{-5}$ was used to identify homologs of switchgrass gene models in the selected genomes.

Phylogenetic analysis

Based on our BLAST search results, we extracted GT2 family protein sequences from *Arabidopsis*, poplar, rice, *Sorghum*, maize, *Brachypodium* and foxtail millet using TAIR V9 and Phytozome V6.0. The GT2 protein sequences were aligned using ClustalX (http://www.clustal.org/) [65] with gap opening/extend penalty 15/0.1 for pairwise alignments and 15/0.2 for multiple alignments. The phylogenetic tree was constructed using the neighbor joining method [66] with 500 bootstrap replicates using clustalX. We used Figtree (http://tree.bio.ed.ac.uk/software/figtree/) to visualize the phylogenetic tree and exported the graphics for formatting in Adobe Illustrator CS4.

Meme motif analysis

The conserved motifs in GT2 family proteins were identified using MEME Suite release 4.9.1 [67] with the following parameters: optimum width 5-50 amino acids, any number of repetitions of a motif and maximum number of motifs set to 15. All the motifs were checked in the NCBI's conserved domain database for significance.

Plant materials, RNA isolation and qPCR analysis

We used greenhouse grown ten-week-old switchgrass cv. Alamo plants for gene expression analysis. We harvested two biological replicates of ~100 mg from six plant organs including root, internode, node, young leaf, mature leaf and panicle. The tissues were frozen in liquid nitrogen and stored at -70ºC until use. RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer's instructions. Total RNA was treated with DNase I and purified with NucleoSpin RNA II kit (Macherey-Nagel). RNA quantity was measured on a Nanodrop (ND-1000 spectrophotometer) from Thermo Scientific. We used Superscript® VILO™ cDNA

synthesis kit (Invitrogen) for cDNA synthesis following the manufacturer's instructions. After quantification and dilution, about 100 ng of cDNA was used as the template for each qPCR reaction using SsoFast EvaGreen Supermix (Bio-Rad) on a Bio-Rad CF96 Real-Time system coupled to a C1000 Thermal Cycler (Bio-Rad). For gene expression normalization, the switchgrass gene encoding for elongation factor was used as an internal control. Relative transcript abundance was measured using comparative delta Ct method [68]. qPCR primer information is provided in supplementary table 8.

Expression data for switchgrass genes was downloaded from the switchgrass functional genomics server (http://switchgrassgenomics.noble.org/index.php). The homologous rice genes were extracted by doing blast search of switchgrass sequences in rice genome annotation project database (http://rice.plantbiology.msu.edu/analyses_search_blast.shtml). The metadata for rice genes was downloaded from the rice array database (http://www.ricearray.org/expression/meta_analysis.shtml).

Acknowledgements

This work was primarily supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 to U.S. Department of Energy Joint Genome Institute and Office of Biological and Environmental Research of the U.S. DOE contract no. DE-AC02-05CH11231 to the Joint BioEnergy Institute. Partial funding for this research was provided by the NSF CREATE-IGERT program at UC Davis (Award Number DGE-0653984).

Figures Legends:

Figure 1. Localization of rice glycosyltransferse (GT), glycoside hydrolase (GH), kinase and ethylene-responsive transcription factor (ERF) family genes and the corresponding switchgrass BAC clones on rice chromosomes

Rice genes belonging to GT, GH, Kinase and ERF families were localized on rice chromosomes based on the chromosomal coordinates given in the rice annotation project database (http://rice.plantbiology.msu.edu/). Colored horizontal bars represent rice genes. Orange, RD kinases; Pink, non-RD kinases; Green, glycosyltransferase; Blue, glycoside hydrolase. The clone IDs of switchgrass BAC clones selected for the targeted genomic regions of rice have been marked. The BAC clones which did not exhibit significant synteny with a corresponding region in rice are marked with a red asterisk.

Figure. 2. Gene ontology analysis of the 3948 genes identified in this study.

Gene ontology terms for the 3948 genes identified in this study were used to generate a 3D pie charts for a) Biological functions; b) Molecular functions and c) Cellular locations categories. Notable categories targeted in this study have been sliced out.

Figure. 3. Phylogenetic analysis of GT2 family genes.

Protein sequences of 506 GT2 family genes from switchgrass, rice, maize, foxtail millet, *Sorghum*, *Brachypodium*, *Arabidopsis* and poplar were used for constructing a phylogenetic tree. The red lines represent proteins from grass genomes, whereas, the green lines represent the eudicot proteins. The accession numbers of GT2 genes from rice (red), *Arabidopsis* (black) and switchgrass (blue) have been provided. The monocot-specific clades are given red background, whereas, eudicot specific clades are shaded green. The names of all clades are given, CesA: Cellulose synthases, Csl: Cellulose synthase-like, Dol: Dolichyl-phosphate beta-glucosyltransferases.

Figure. 4. Microarray-based comparative expression analysis of rice and switchgrass glycosyltransferase 2 family genes

The heat map displays microarray-based expression data of homologous genes of rice and switchgrass in six developmental tissues. The given names of rice genes and accession numbers of homologous switchgrass genes are provided on the left. The numbers on the right indicate Pearson's correlation coefficient values between expression patterns of homologous genes of rice and switchgrass. The developmental stages are given at the top. The expression data for root, leaf blade, leaf sheath and stem tissue in switchgrass were taken at the elongation stage. The inflorescence 1 stage represents inflorescence meristem and inflorescence 2 represents inflorescence with floret development. The metadata for comparable tissues in rice, based on the Agilent 44K platform, was downloaded from the Rice Oligonucleotide Array Database, www.ricearray.org. The legend for relative expression values for both color schemes is given at the base of the heat map. The red-black-green color scheme represents expression of switchgrass genes with red indicating high expression, black medium expression and green showing low-level expression. The yellow-black-blue color scheme indicates expression patterns of rice genes with yellow indicating high expression, black medium expression and blue low-level expression.

Figure. 5. QPCR-based expression analysis of representative GT2 family genes of switchgrass.

The bar graphs present the expression patterns of 24 selected GT2 family genes of switchgrass in six developmental tissues using qPCR. The values on the Y-axis indicate relative expression. Developmental stages are given on the X-axis as follows: R Root, P Panicle, ML Mature leaf, YL young leaf, In Internode and N Node. The error bars represent standard error for two biological replicates.

Table 1. List of switchgrass genes homologous to rice genes involved in bioenergy-relevant traits.

Supplementary Files:

Supplementary Figure 1. Flow chart showing the complete BAC library screening strategy used in this study.

Supplementary Figure 2. Dendrogram showing phylogenetic relationship between GT2 genes from switchgrass, rice, maize, *Sorghum*, foxtail millet, *Brachypodium*, poplar and *Arabidopsis*.

Supplementary Figure 3. Results of MEME motif analysis of protein sequences used for constructing the phylogenetic tree.

Suppl Table 1. BAC statistics

Suppl Table 2. List of genes annotated from switchgrass BAC clones.

Suppl Table 3. List of Simple Sequence Repeats identified from full-length BAC sequences.

Suppl Table 4. List of known repetitive elements identified from full-length BAC sequences.

Suppl Table 5. List of GT, GH, kinase and ERF family genes annotated from switchgrass BAC clones.

Suppl Table 6. Homologs of genes annotated from switchgrass full-length BAC sequences from five grass genomes and *Arabidopsis*.

Supp Table 7. Consensus sequences of MEME motifs identified from GT2 proteins.

Supp Table 8. List of qPCR primers designed from GT2 genes.

Suppl File 1. Genomic sequences of switchgrass genes annotated from full-length sequences of BAC clones.

Suppl File 2. cDNA sequences of switchgrass genes annotated from full-length BAC clones.

Suppl File 3. Peptide sequences of switchgrass proteins annotated from full-length BAC clones.

Suppl File 4. gff file – annotation information.

Suppl File 5. Sequences of GT2 proteins from switchgrass, foxtail millet, *Sorghum*, maize, rice, *Arabidopsis* and poplar used for phylogenetic analysis.

References

1. Sanderson MA, Reed RL, McLaughlin SB, Wullschleger SD, Congcr BV, et al. (1996) Switchgrass as a sutainable bioenergy crop. Bioresource Technology 56: 83-93.
2. Barney JN, Mann JJ, Kyser GB, Blumwald E, Deynze AV, et al. (2009) Tolerance of switchgrass to extreme soil moisture stress: Ecological implications. Plant Science 177: 724-732.
3. Fike J, Parrish D, Wolf D, Balasko J, Green Jr. J, et al. (2006) Long-term yield potential of switchgrass-for-biofuel systems. Biomass Bioenergy 30: 198-206.
4. McLaughlin SB, Kszos LA (2005) Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. Biomass and Bioenergy 28: 515-535.

5. Carroll A, Somerville C (2009) Cellulosic biofuels. Annu Rev Plant Biol 60: 165-182.
6. Kristensen JB, Felby C, Jorgensen H (2009) Yield-determining factors in high-solids enzymatic hydrolysis of lignocellulose. Biotechnol Biofuels 2.
7. Sims RE, Mabee W, Saddler JN, Taylor M (2010) An overview of second generation biofuel technologies. Bioresour Technol 101: 1570-1580.
8. van der Weijde T, Alvim Kamei CL, Torres AF, Vermerris W, Dolstra O, et al. (2013) The potential of C4 grasses for cellulosic biofuel production. Front Plant Sci 4: 107.
9. Cannella D, Jorgensen H (2014) Do New Cellulolytic Enzyme Preparations Affect the Industrial Strategies for High Solids Lignocellulosic Ethanol Production? Biotechnology and Bioengineering 111: 59-68.
10. Davidson S (2008) Sustainable Bioenergy: Genomics and biofuels development. Nature Education 1: 175.
11. Johnson JM-F, Coleman MD, Gesch R, Jaradat A, Mitchell R, et al. (2007) Biomass-Bioenergy crops in the United States: A changing paradigm. The Americas Journal of Plant Sciences and Biotechnology 1: 1-28.
12. Nageswara-Rao M, Soneji JR, Kwit C, Stewart CN, Jr. (2013) Advances in biotechnology and genomics of switchgrass. Biotechnol Biofuels 6: 77.
13. Casler MD, Tobias CM, Kaeppler SM, Buell CR, Wang ZY, et al. (2011) The Switchgrass Genome: Tools and Strategies. The Plant Genome 4: 273-282.
14. Cosgrove DJ (2005) Growth of the plant cell wall. Nature Reviews Molecular Cell Biology 6: 850-861.
15. Kohorn BD, Kohorn SL (2012) The cell wall-associated kinases, WAKs, as pectin receptors. Front Plant Sci 3: 88.
16. Sharma R, De Vleesschauwer D, Sharma MK, Ronald PC (2013) Recent Advances in Dissecting Stress-Regulatory Crosstalk in Rice. Molecular Plant 6: 250-260.
17. Sharma R, Tan F, Jung KH, Sharma MK, Peng ZH, et al. (2011) Transcriptional dynamics during cell wall removal and regeneration reveals key genes involved in cell wall development in rice. Plant Molecular Biology 77: 391-406.
18. Dardick C, Chen J, Richter T, Ouyang S, Ronald P (2007) The rice kinase database. A phylogenomic database for the rice kinome. Plant Physiol 143: 579-586.
19. Jung KH, Cao P, Seo YS, Dardick C, Ronald PC (2010) The Rice Kinase Phylogenomics Database: a guide for systematic analysis of the rice kinase super-family. Trends Plant Sci 15: 595-599.
20. Dardick C, Ronald P (2006) Plant and animal pathogen recognition receptors signal through non-RD kinases. PLoS Pathog 2: e2.
21. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms 1819: 86-96.
22. Sharma MK, Kumar R, Solanke AU, Sharma R, Tyagi AK, et al. (2010) Identification, phylogeny, and transcript profiling of ERF family genes during development and abiotic stress treatments in tomato. Mol Genet Genomics 284: 455-475.
23. Xu ZS, Chen M, Li LC, Ma YZ (2011) Functions and Application of the AP2/ERF Transcription Factor Family in Crop Improvement. Journal of Integrative Plant Biology 53: 570-585.

24. Ariyadasa R, Stein N (2012) Advances in BAC-based physical mapping and map integration strategies in plants. J Biomed Biotechnol 2012: 184854.
25. Campbell TN, Choy FY (2002) Approaches to library screening. J Mol Microbiol Biotechnol 4: 551-554.
26. Sharma MK, Sharma R, Cao PJ, Jenkins J, Bartley LE, et al. (2012) A Genome-Wide Survey of Switchgrass Genome Structure and Organization. Plos One 7.
27. Clark L, Carbon J (1976) A colony bank containing synthetic Col E1 hybrids representative of the entire *E. coli* genome. Cell 9: 91-99.
28. Bouzidi MF, Franchel J, Tao Q, Stormo K, Nicolas MP, et al. (2006) A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions. Theor Appl Genet 113: 81-89.
29. Vu GT, Caligari PD, Wilkinson MJ (2010) A simple, high throughput method to locate single copy sequences from Bacterial Artificial Chromosome (BAC) libraries using High Resolution Melt analysis. BMC Genomics 11: 301.
30. Seo YS, Chern M, Bartley LE, Han M, Jung KH, et al. (2011) Towards establishment of a rice stress response interactome. PLoS Genet 7: e1002020.
31. Okada M, Lanzatella C, Saha MC, Bouton J, Wu R, et al. (2010) Complete Switchgrass Genetic Maps Reveal Subgenome Collinearity, Preferential Pairing, and Multilocus Interactions. Genetics doi:10.1534/genetics.110.113910
32. Pandey G, Misra G, Kumari K, Gupta S, Parida SK, et al. (2013) Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [Setaria italica (L.)]. DNA Res 20: 197-207.
33. Heredia A, Jimenez A, Guillen R (1995) Composition of Plant-Cell Walls. Zeitschrift Fur Lebensmittel-Untersuchung Und-Forschung 200: 24-31.
34. Vogel J (2008) Unique aspects of the grass cell wall. Current Opinion in Plant Biology 11: 301-307.
35. Carroll A, Specht CD (2011) Understanding Plant Cellulose Synthases through a Comprehensive Investigation of the Cellulose Synthase Family Sequences. Front Plant Sci 2: 5.
36. Mellerowicz EJ, Baucher M, Sundberg B, Boerjan W (2001) Unravelling cell wall formation in the woody dicot stem. Plant Molecular Biology 47: 239-274.
37. Yin Y, Huang J, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and algae. Bmc Plant Biology 9: 99.
38. Wang YW, Samuels TD, Wu YQ (2010) Development of 1,030 genomic SSR markers in switchgrass. Theor Appl Genet Oct 27. [Epub ahead of print].
39. Cao P, Jung KH, Choi D, Hwang D, Zhu J, et al. (2012) The Rice Oligonucleotide Array Database: an atlas of rice gene expression. Rice (N Y) 5: 17.
40. Zhang JY, Lee YC, Torres-Jerez I, Wang M, Yin Y, et al. (2013) Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (Panicum virgatum L.). Plant J 74: 160-173.
41. Gabaldon T, Koonin EV (2013) Functional and evolutionary implications of gene orthology. Nat Rev Genet 14: 360-366.
42. Wang LQ, Guo K, Li Y, Tu YY, Hu HZ, et al. (2010) Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. Bmc Plant Biology 10.
43. Endler A, Persson S (2011) Cellulose synthases and synthesis in Arabidopsis. Mol Plant 4: 199-211.

44. Kumar M, Turner S (2014) Plant cellulose synthesis: CESA proteins crossing kingdoms. Phytochemistry.
45. Li M, Xiong G, Li R, Cui J, Tang D, et al. (2009) Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. Plant J 60: 1055-1069.
46. Yoshikawa T, Eiguchi M, Hibara KI, Ito JI, Nagato Y (2013) Rice SLENDER LEAF 1 gene encodes cellulose synthase-like D4 and is specifically expressed in M-phase cells to regulate cell proliferation. Journal of Experimental Botany 64: 2049-2061.
47. Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, et al. (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. Science 311: 1940-1942.
48. Vega-Sanchez ME, Verhertbruggen Y, Christensen U, Chen XW, Sharma V, et al. (2012) Loss of cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. Plant Physiology 159: 56-69.
49. Xie G, Yang B, Xu Z, Li F, Guo K, et al. (2013) Global identification of multiple OsGH9 family members and their involvement in cellulose crystallinity modification in rice. PLoS One 8: e50171.
50. Chen W, VanOpdorp N, Fitzl D, Tewari J, Friedemann P, et al. (2012) Transposon insertion in a cinnamyl alcohol dehydrogenase gene is responsible for a brown midrib1 mutation in maize. Plant Mol Biol 80: 289-297.
51. Dalmais M, Antelme S, Ho-Yue-Kuang S, Wang Y, Darracq O, et al. (2013) A TILLING Platform for Functional Genomics in. PLoS One 8: e65503.
52. Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, et al. (2011) Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proc Natl Acad Sci U S A 108: 3803-3808.
53. Saathoff AJ, Sarath G, Chow EK, Dien BS, Tobias CM (2011) Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. PLoS One 6: e16416.
54. Eom JS, Cho JI, Reinders A, Lee SW, Yoo Y, et al. (2011) Impaired function of the tonoplast-localized sucrose transporter in rice, OsSUT2, limits the transport of vacuolar reserve sucrose and affects plant growth. Plant Physiol 157: 109-119.
55. Brown DM, Zhang Z, Stephens E, Dupree P, Turner SR (2009) Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. Plant J 57: 732-746.
56. Chiniquy D, Varanasi P, Oh T, Harholt J, Katnelson J, et al. (2013) Three novel rice genes closely related to the *Arabidopsis IRX9*, *IRX9L*, and *IRX14* genes and their roles in xylan biosynthesis. Front Plant Sci 4: 83.
57. Dai X, You C, Chen G, Li X, Zhang Q, et al. (2011) OsBC1L4 encodes a COBRA-like protein that affects cellulose synthesis in rice. Plant Mol Biol 75: 333-345.
58. Zhou Y, Li S, Qian Q, Zeng D, Zhang M, et al. (2009) BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (Oryza sativa L.). Plant J 57: 446-462.
59. Jensen E, Robson P, Norris J, Cookson A, Farrar K, et al. (2013) Flowering induction in the bioenergy grass Miscanthus sacchariflorus is a quantitative short-day response, whilst delayed flowering under long days increases biomass accumulation. J Exp Bot 64: 541-552.

60. Salehi H, Ransom CB, Oraby HF, Seddighi Z, Sticklen MB (2005) Delay in flowering and increase in biomass of transgenic tobacco expressing the Arabidopsis floral repressor gene FLOWERING LOCUS C. J Plant Physiol 162: 711-717.
61. Cantu D, Yang B, Ruan R, Li K, Menzo V, et al. (2013) Comparative analysis of protein-protein interactions in the defense response of rice and wheat. BMC Genomics 14: 166.
62. Kolpakov R, Bana G, Kucherov G (2003) mreps: Efficient and flexible detection of tandem repeats in DNA. Nucleic acids research 31: 3672-3678.
63. Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] Chapter 4: Unit 4 10.
64. Smit A, Hubley R, P. G (2010) RepeatMasker Open-3.0. 1996-2010.
65. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876-4882.
66. Saitou N, Nei M (1987) The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. Molecular Biology and Evolution 4: 406-425.
67. Bailey TL, Gribskov M (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proceedings of the Second International conference on intelligent System for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California.
68. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods 25: 402-408.